# Point cloud subjective evaluation methodology based on 2D rendering

Evangelos Alexiou*, Marco V. Bernardo†, Luis A. da Silva Cruz‡, Lovorka Gotal Dmitrović§, Carlos Duarte‡, Emil Dumić§, Touradj Ebrahimi*, Dragan Matković§, Manuela Pereira†, Antonio Pinheiro† and Athanasios Skodras¶
*Multimedia Signal Processing Group, École Polytechnique Fédérale de Lausanne, Switzerland
†Institute of Telecommunications, University of Beira Interior, Portugal
‡Department of Electrical and Computer Engineering, University of Coimbra and Institute of Telecommunications, Portugal
§Department of Electrical Engineering, University North, Croatia
¶Department of Electrical and Computer Engineering, University of Patras, Greece

***Abstract*—Point clouds are one of the most promising technologies for 3D content representation. In this paper, we describe a study on quality assessment of point clouds, degraded by octree-based compression on different levels. The test contents were displayed using Screened Poisson surface reconstruction, without including any textural information, and they were rated by subjects in a passive way, using a 2D image sequence. Subjective evaluations were performed in five independent laboratories in different countries, with the inter-laboratory correlation analysis showing no statistical differences, despite the different equipment employed. Benchmarking results reveal that the state-of-the-art point cloud objective metrics are not able to accurately predict the expected visual quality of such test contents. Moreover, the subjective scores collected from this experiment were found to be poorly correlated with subjective scores obtained from another test involving visualization of raw point clouds. These results suggest the need for further investigations on adequate point cloud representations and objective quality assessment tools.**

***Index Terms*—Quality Assessment, Point Cloud, Quality Metrics**

## I. Introduction

Point Clouds (PCs) are emerging as a viable solution to efficiently represent 3D geometric and visual information. This trend is assisted by the current availability of low-cost, high-performance depth sensors, and the integration of powerful signal processing units in typical electronic devices. In many applications, PC data are not meant to be displayed directly, rather processed to extract information about the imaged scenery; an indicative example is the usage of Light Detection And Ranging (LiDAR) technologies in autonomous driving. However, in other application scenarios, such as in entertainment industry for instance, humans directly consume PC data. Since the large volumes of such contents require the use of compression schemes to manage the sheer scale of this data size, efficient ways to represent these data are needed, with and without information loss.

As in any other type of imaging, compression of PCs requires efficient objective distortion measures that are well matched to subjective opinion of human observers. Several objective quality metrics have been proposed in the literature and their correlation with subjective quality scores has already been investigated. However, in most such studies, PCs were displayed as a set of points without any intervening surface reconstruction algorithm; the latter reflects a rather common way to consume 3D contents nowadays. In particular, in [1] subjective assessment of raw PCs degraded by geometry and color, after applying uniform noise was performed. In [2], the performance of a deployed codec for real-time dynamic PC sequence was subjectively assessed in a realistic 3D tele-immersive system. The users, represented as 3D avatars (synthetic content) and/or 3D PC (naturalistic content), were able to navigate and interact with the assets of a virtual 3D room. Different aspects of the quality were assessed, including the overall quality of the 3D human rendition, the quality of colors, and the level of immersiveness, among others. However, in the aforementioned studies no correlation between subjective and objective scores was reported.

In [3], PC denoising algorithms were subjectively evaluated and the test contents were visualized after applying the Screened Poisson surface reconstruction [4]. A passive assessment was adopted and 2D video sequences were formed, after capturing the resulting mesh objects from different viewpoints by vertical and horizontal rotation. However, the impact of visualizing reconstructed meshes instead of raw PCs was not investigated. In [5], subjective quality assessment of colored PCs was conducted, subject to simple octree and graph-based encoding algorithms. To render the PC data, primitive cubes were used. The resulted test contents were captured from different viewing angles with the virtual camera following a spiral path. The subjects visualized animated 2D video sequences to provide their scores. In [6] and [7], an interactive approach to subjectively assess geometry-only PCs in a desktop setting was proposed, using the Double-Stimulus Impairment Scale (DSIS) and the Absolute Category Rating (ACR) evaluation methodologies, respectively; the latter study also investigated the correlation between them. The contents under evaluation were degraded by introducing Gaussian noise, and octree-based compression, while the PCs were rendered as a set of points. In [8], a subjective methodology for PCs using head mounted displays in an augmented reality scenario was proposed. The subjects visualized raw PCs and their interaction with the 3D models was performed by physical movements in the real world environment.

In this article we describe a set of PC evaluations, where raw PCs are first converted to watertight 3D objects, which are then subjectively assessed. The tests are conducted at five different research laboratories with different display equipment and correlation results are reported. Performance indexes after benchmarking of the state-of-the-art objective metrics are also provided. Finally, we present comparison results between subjective scores obtained in two experiments involving different data representations, namely, visualization of (a) reconstructed objects as polygonal meshes and (b) raw PCs as set of points.

## II. Subjective Assessment

In this section the preparation of test contents is described, followed by the design of the subjective evaluations.

### A. Content Preparation

A dataset of 7 geometry-only PCs was used. In particular, *bunny* and *dragon* were selected from the Stanford 3D Scanning repository. *Egyptian_mask* is a content used in the recent activities of the MPEG standardization committee [9], and *vase* is an object captured by Intel RealSense R200 in [6]. *Cube* and *sphere* were synthesized using corresponding mathematical formulas, while *torus* was artificially produced in MeshLab. To ensure that the number of points of every model is in the same order of magnitude, corresponding releases (i.e., *dragon_vrip_res3*) were selected, or sub-sampled versions (i.e., *vase* and *egyptian_mask*) were generated without modifying the original coordinates, thus, maintaining the original geometric structure of the test contents.

The original PCs were compressed after applying octree pruning, as described in [7], in the Point Cloud Library (PCL) v1.8.0 [10]. In particular, the contents were enclosed in an octee structure. By modifying the size of the leaf nodes, which is referred as Level of Details (LoD), the resolution of the content is correspondingly adjusted. For instance, after increasing the LoD, the number of points of the compressed object naturally decreases. Considering that the octree is the basis for the majority of PC compression algorithms, this is a simplified approach to produce visible artifacts after octree-based encoding. To account for a wide range of visible distortions, the target percentages (p) of remaining points after octree pruning were selected as: $90\%$, $70\%$, $50\%$ and $30\%$, allowing a deviation of $\pm 2\%$. For *torus*, an additional version with $98\%$ of points was also prepared to be used in the training. The number of points for every reference and distorted content, along with the LoD values that were used, can be found in Table I.

The original raw PCs were initially scaled to fit in a bounding box of size $1$ and translated at the origin $(0, 0, 0)$. Then, the distorted versions were produced following the aforementioned procedure. The Screened Poisson surface reconstruction algorithm [4] was selected to be used in order to consume the objects. The CloudCompare implementation was employed setting an octree depth of $8$ and default parameters. This method is popular due to (i) availability of open source software, (ii) guaranteed generation of watertight objects, (iii) adjustable complexity, as a function of the octree depth, and (iv) reproducibility of the generated meshes; the latter not being a given feature in reconstruction techniques. To apply this algorithm, normal vectors should exist along with the coordinates of every content. As normal vectors were absent, the estimation was performed in CloudCompare with default settings, i.e., the radius to identify nearest neighbors was selected automatically and a plane was used as the local surface model. Then, on the same tool, the normals were oriented using a Minimum Spanning Tree of 6 nearest neighbors.

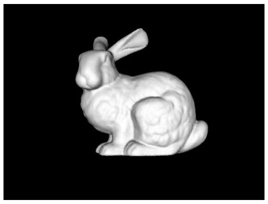
TABLE I
Parameters used for octree pruning.

| | LoD | Number of points | Actual percentage | Target percentage |
|---|---|---|---|---|
| *bunny* | - | 35947 | 100.00% | 100 |
| | 0.007 | 32957 | 91.68% | 90 |
| | 0.010 | 25209 | 70.13% | 70 |
| | 0.012 | 17763 | 49.41% | 50 |
| | 0.016 | 10870 | 30.24% | 30 |
| *cube* | - | 30246 | 100.00% | 100 |
| | 0.015 | 27541 | 91.06% | 90 |
| | 0.017 | 20888 | 69.06% | 70 |
| | 0.020 | 15002 | 49.60% | 50 |
| | 0.025 | 9602 | 31.75% | 30 |
| *dragon* | - | 22998 | 100.00% | 100 |
| | 0.008 | 20847 | 90.65% | 90 |
| | 0.010 | 16487 | 71.69% | 70 |
| | 0.013 | 11539 | 50.17% | 50 |
| | 0.017 | 7026 | 30.55% | 30 |
| *egyptian_mask* | - | 31601 | 100.00% | 100 |
| | 0.008 | 28393 | 89.85% | 90 |
| | 0.010 | 22061 | 69.81% | 70 |
| | 0.013 | 15790 | 49.97% | 50 |
| | 0.017 | 9466 | 29.96% | 30 |
| *sphere* | - | 30135 | 100.00% | 100 |
| | 0.004 | 27298 | 90.59% | 90 |
| | 0.011 | 21100 | 70.02% | 70 |
| | 0.015 | 15168 | 50.33% | 50 |
| | 0.020 | 8977 | 29.79% | 30 |
| *vase* | - | 36022 | 100.00% | 100 |
| | 0.007 | 32454 | 90.10% | 90 |
| | 0.009 | 25217 | 70.00% | 70 |
| | 0.011 | 17963 | 49.87% | 50 |
| | 0.015 | 10693 | 29.69% | 30 |
| *torus* (reference) | - | 31250 | 100.00% | 100 |
| | 0.005 | 30566 | 97.81% | 98 |
| | 0.007 | 27968 | 89.50% | 90 |
| | 0.010 | 21901 | 70.08% | 70 |
| | 0.012 | 15715 | 50.29% | 50 |
| | 0.017 | 9539 | 30.53% | 30 |

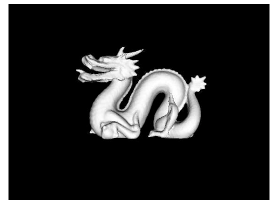
The PCL visualizer was used to render the contents, by setting the background color to black, and using default lighting conditions and flat shading. The models were placed at the origin of the virtual environment and a fixed distance from the camera was set to avoid changes of the model's size that may be perceived as the camera is circularly moving around it. The camera rotated around the horizontal and, then, around the vertical axis of the center of the object in steps of $1°$. In every step, a still frame was captured, leading to a total of $720$ frames. The still images were then losslessly compressed with an H.264/AVC encoder, producing an animated video of 30 fps with a total duration of $24$ seconds. The reference meshes of every selected content are shown in Figure 1, while in Figure 2 the *bunny* is presented for every degradation level.

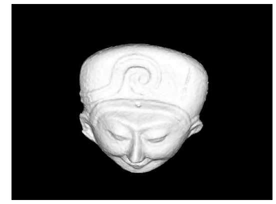
### B. Evaluation Methodology

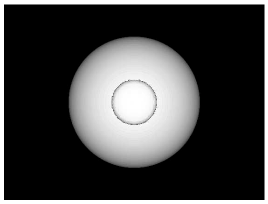
The subjective experiments were conducted in $5$ laboratories: École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; University of Beira Interior (UBI), Covilhã, Portugal; University of Coimbra (UC), Coimbra,

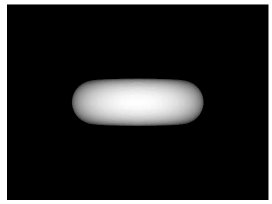

(a) *bunny*


(b) *cube*


(c) *dragon*

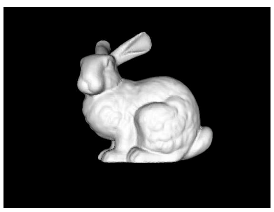

(d) *egyptian mask*

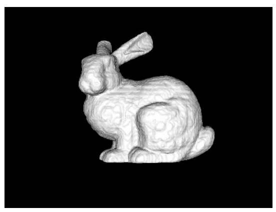

(e) *sphere*

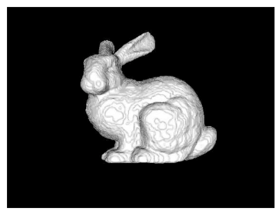

(f) *torus*

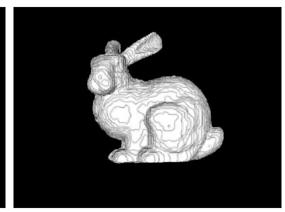

(g) *vase*

Fig. 1. Frontal view of each reference mesh.


(a) 100%


(b) 90%


(c) 70%


(d) 50%

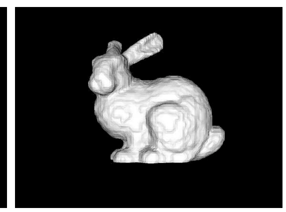

(e) 30%

Fig. 2. Frontal view of *bunny* for every target percentage.

TABLE II
EQUIPMENT AND SUBJECTS INFORMATION PER LABORATORY.

| | | **EPFL** | **UBI** | **UC** | **UNIN** | **UP** |
|---|---|---|---|---|---|---|
| *Equipment* | Monitor | Apple Cinema M9179LL/A | ASUS PB287Q | Sony KD-49X8005C | Sony KD55x8505 | Dimenco DM504MA5 |
| | Inches | 30” | 28” | 49” | 55” | 50” |
| | Resolution | 2560x1600 | 3840x2160 | 3840x2160 | 3840x2160 | 1920x1080 |
| | View Distance | 0.7 m (FV) | 1.5 m (FV) | 1.8 m (FV $\pm$30 cm) | 1.5 m (FV) | 1.5 m (FV) |
| *Subject Info* | Males | 11 | 17 | 9 | 14 | 30 |
| | Females | 9 | 5 | 11 | 6 | 14 |
| | Overall | 20 | 22 | 20 | 20 | 44 |
| | Year span | 21-37 | 21-50 | 21-54 | 19-57 | 19-59 |
| | Average age | 28.88 | 30.59 | 29.45 | 26.45 | 23.32 |
| | Median age | 28.39 | 28 | 23 | 21.5 | 22 |
| | Outliers | 1 | 0 | 0 | 1 | 6 |

Portugal; University North (UNIN), Varaždin, Croatia and Univeristy of Patras (UP), Patras, Greece. The conditions of every test environment were adjusted to follow the ITU-R Recommendation BT.500-13 [11], while the equipment used per laboratory can be found in Table II. A passive subjective methodology was applied, with the subjects visualizing the generated 2D viseo sequences in the mpv video player, and providing their scores using a customized interface, either during or after the completion of the playback animation.

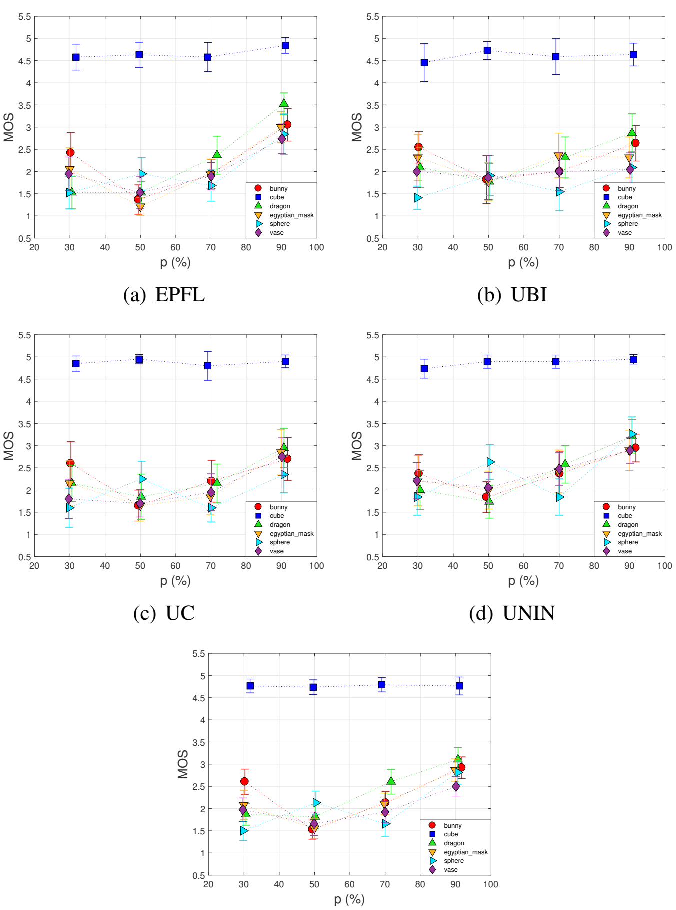
The DSIS simultaneous test method was adopted with a 5-level impairment scale, including a hidden reference for sanity check. Thus, both the reference and the degraded stimuli were simultaneously shown to the observer, side-by-side, and every subject rated the visual quality of the processed with respect to the reference stimulus. To avoid biases, in half of the individual evaluations, the reference was placed on the right and the degraded content on the left side of the screen, and vice-versa for the rest of the evaluations. Particular care was also given to avoid displaying the same model consecutively.

A free viewing (FV) scenario was adopted for the assessment; that is, after the initial position, which was defined in every laboratory and reported in Table II, every subject was free to move closer or further from the screen during the evaluation. This is because different objects could be perceived of different volume. For instance, from a fixed distance between the observer and the screen, the *dragon* is perceived smaller with respect to the *sphere*, due to the different ratio between height and length. At the beginning of each individual evaluation, a training session took place, in order to familiarize the subjects with the artifacts under assessment. The *torus* content was selected for this purpose and, hence, it was excluded from the actual subjective tests. The training was performed using 3 animated video sequences that represented 3 different levels of degradation in order to indicatively illustrate the range of visible distortions.

An overall of 30 scores were obtained per evaluation session, considering that each subject assessed 6 test contents degraded in 4 distinct levels along with the hidden references. An outlier detection algorithm based on ITU-R Recommendation BT.500-13 [11] was applied to the collected scores, and the ratings of the identified outliers were discarded. Then, the mean opinion scores (MOS) and the 95% Confidence Intervals (CIs), assuming a Student's t-distribution were computed. In Table II, equipment details, observer information and the number of outliers per test laboratory are reported.

## III. Results

The subjective results of the 6 meshes after outlier detection are shown in Figure 3, with the caption of each sub-figure indicating the laboratory from which they were collected.




(a) EPFL (b) UBI

(c) UC (d) UNIN

(e) UP

Fig. 3. Subjective scores against degradation values per laboratory.

Notably, it can be observed that the MOS for *cube* remains high, independently of the level of distortion. For the other meshes the MOS is increasing as the target percentage is increasing, with the exception of the lowest degradation level, where the MOS is stable or even slightly higher. This can be explained by the smoother functions that are used by the reconstruction algorithm to produce the surfaces of the mesh, due to the vast reduction of the number of points for these distorted contents. An example can be viewed in Figure 2, where the impairment of the content of Figure 2 (e) is less annoying when compared to the content of Figure 2 (d).

In the following sections we provide a subset of our results. The reader can access the complete results through the URL: https://mmspg.epfl.ch/reconstructed-point-clouds-results.

### A. Correlation between Subjective and Objective Scores

The subjective scores were correlated with the state-of-the-art PC objective metrics. The point-to-point (p2point) and point-to-plane (p2plane) metrics were used, as implemented in the software v. 0.11 described in [12], with the Mean-Squared-Error (MSE) and the Hausdorff distance accounting for the geometric errors. The corresponding PSNR values, as defined in [12], were also computed using a factor of 1 in the numerator, leading to a total of 8 objective metrics.

For benchmarking of the objective quality assessment tools, typically, the subjective scores are considered as the ground truth. A predicted MOS for a particular distorted content is obtained after applying a fitting function between the subjective MOS and the corresponding objective scores. Based on the Recommendation ITU-T P.1401 [13], the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE) and the Outlier Ratio (OR) were computed between the subjective and the predicted MOS, to account for linearity, monotonicity, accuracy and consistency. Among various functions, the cubic fitting was selected as it provides the best fitting results.

The objective scores were calculated on (a) the raw PCs, and (b) the PCs obtained after the surface reconstruction, given that a mesh consists of a set of points in conjunction with a set of associated faces. To compute the p2plane metrics on the raw PCs, we used the estimated normal vectors as described in Section II-A. To compute the p2plane metrics on the PCs after surface reconstruction, we used the normal vectors that are naturally produced after the generation of the mesh. Due to the outlier behavior of the *cube*, shown in Figure 3, the performance indexes are calculated after including and excluding the scores of this content. In Table III, we present the best-performing objective metric for every case. It is noted that, the correlation is generally poor and better performance is achieved when *cube* is not considered. Moreover, the performance indexes worsen when the objective scores are estimated on the PCs after surface reconstruction. As can be seen, the best correlation is achieved for the subjective scores of UNIN after using the p2point metric with Hausdorff, for raw PCs before surface reconstruction and by excluding the scores of *cube*.

TABLE III
BENCHMARKING RESULTS FOR THE BEST-PERFORMING OBJECTIVE METRIC PER LABORATORY.

**Before** surface reconstruction

| | | Objective metric | PCC | SROCC | RMSE | OR |
|---|---|---|---|---|---|---|
| **EPFL** | With *cube* | p2plane-MSE | 0.584 | 0.023 | 0.922 | 0.750 |
| | Without *cube* | p2point-Hausdorff | 0.740 | 0.591 | 0.432 | 0.400 |
| **UBI** | With *cube* | p2plane-MSE | 0.680 | 0.163 | 0.744 | 0.500 |
| | Without *cube* | p2plane-Hausdorff | 0.532 | 0.392 | 0.306 | **0.200** |
| **UC** | With *cube* | p2plane-MSE | 0.678 | 0.067 | 0.821 | 0.625 |
| | Without *cube* | p2point-MSE | 0.617 | 0.528 | 0.347 | 0.300 |
| **UNIN** | With *cube* | p2plane-MSE | 0.622 | -0.066 | 0.811 | 0.708 |
| | Without *cube* | p2point-Hausdorff | **0.834** | **0.727** | **0.258** | 0.250 |
| **UP** | With *cube* | p2plane-MSE | 0.640 | 0.015 | 0.839 | 0.833 |
| | Without *cube* | p2point-Hausdorff | 0.722 | 0.607 | 0.352 | 0.550 |

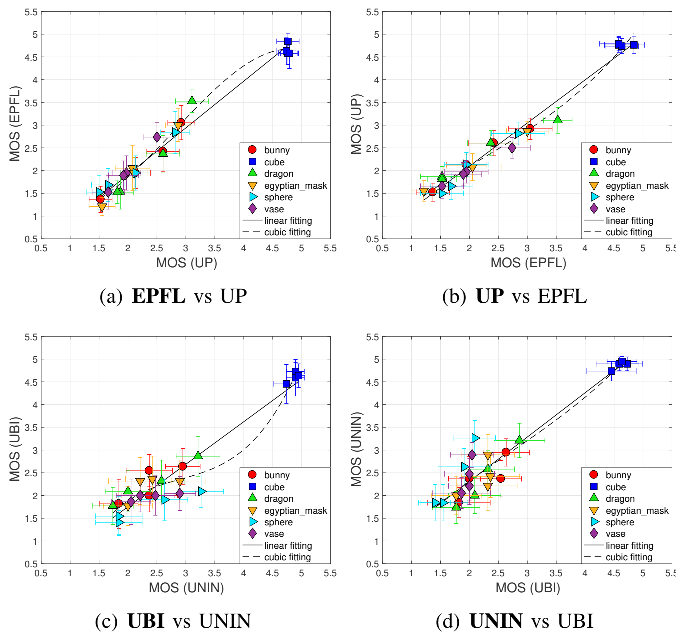
**After** surface reconstruction

| | | Objective metric | PCC | SROCC | RMSE | OR |
|---|---|---|---|---|---|---|
| **EPFL** | With *cube* | p2plane-MSE | 0.653 | 0.038 | 0.860 | 0.667 |
| | Without *cube* | p2plane-MSE | 0.488 | 0.627 | 0.561 | 0.600 |
| **UBI** | With *cube* | p2plane-MSE | **0.773** | 0.091 | 0.643 | 0.417 |
| | Without *cube* | p2point-Hausdorff | 0.549 | 0.427 | **0.302** | **0.100** |
| **UC** | With *cube* | p2plane-MSE | 0.771 | 0.089 | 0.711 | 0.667 |
| | Without *cube* | p2point-MSE | 0.389 | 0.380 | 0.406 | 0.350 |
| **UNIN** | With *cube* | p2plane-MSE | 0.706 | -0.016 | 0.733 | 0.667 |
| | Without *cube* | p2plane-MSE | 0.549 | **0.717** | 0.391 | 0.450 |
| **UP** | With *cube* | p2plane-MSE | 0.723 | 0.038 | 0.754 | 0.875 |
| | Without *cube* | p2plane-MSE | 0.436 | 0.631 | 0.458 | 0.650 |

## B. Comparison between Subjective Scores from different Labs

To compare the subjective scores between the participated laboratories, the PCC, SROCC, RMSE and OR correlation coefficients were calculated. Furthermore, the Correct Estimation (CE), Under Estimation (UE) and Over Estimation (OE) percentages were realized, as proposed in the Recommendation ITU-T P.1401 [13], to check for statistically equivalent MOS results. Moreover, the False Ranking (FR), False Differentiation (FD), False Tie (FT) and Correct Decision (CD) percentages were implemented, based on the Recommendation ITU-T J.149 [14], to check for different conclusions on a pair of data points. Since the scores of a particular laboratory cannot be used as the ground truth, for every pair of universities $(A, B)$, the subjective scores of university $A$ are considered as the ground truth while the scores of university $B$ are benchmarked, and vice versa. Every performance index was computed using no fitting, linear fitting and cubic fitting functions. The correlation results were similar in every case, but slightly better with cubic fitting, which is used for the coefficients reported in Table IV. Please notice that not every performance index is presented, to avoid redundancies. In particular, CE remains at $100\%$ in every tested case, except of UBI against UNIN, where an OE of 4% was found for the scores of UNIN without fitting. Furthermore, the FR, the most offensive type of error, is always $0\%$ and, thus, the CD can be obtained by subtracting the sum of FD and FT from $100\%$.

Based on our results, strong correlation can be observed in every case. A remark is that the FT percentages are consistently high when the UP is the ground truth, as the CIs are small due to the high number of involved subjects. A general conclusion is that subjective evaluations of this visual modality do not highly depend on the specifications of the monitor in a desktop setting. Moreover, the free viewing approach further assists to obtain these results. In Figure 4, scatter plots showing the comparison of MOS between two pairs of universities are indicatively presented.




(a) **EPFL** vs UP
(b) **UP** vs EPFL
(c) **UBI** vs UNIN
(d) **UNIN** vs UBI

Fig. 4. No, Linear and Cubic fitting, to evaluate the correlation between pairs of laboratories (Bold text represents the ground truth).

TABLE IV
PERFORMANCE INDEXES TO COMPARE SUBJECTIVE SCORES BETWEEN DIFFERENT LABS (BOLD TEXT REPRESENTS THE GROUND TRUTH).

| | PCC | SROCC | RMSE | OR | FD ( %) | FT (%) |
|---|---|---|---|---|---|---|
| **EPFL** *vs* UBI | 0.947 | 0.890 | 0.365 | 0.333 | 2.17% | 12.68% |
| **UBI** *vs* EPFL | 0.975 | 0.890 | 0.226 | 0.125 | 0.36% | 2.17% |
| **EPFL** *vs* UC | 0.975 | 0.922 | 0.251 | 0.167 | 3.62% | 6.52% |
| **UC** *vs* EPFL | 0.984 | 0.922 | 0.196 | 0.083 | 3.26% | 4.35% |
| **EPFL** *vs* UNIN | 0.979 | 0.927 | 0.232 | 0.125 | 1.45% | 9.42% |
| **UNIN** *vs* EPFL | 0.984 | 0.927 | 0.187 | 0.083 | 2.17% | 3.62% |
| **EPFL** *vs* UP | 0.992 | 0.969 | 0.144 | 0.042 | 7.97% | 1.45% |
| **UP** *vs* EPFL | 0.991 | 0.969 | 0.143 | 0.125 | 1.09% | 9.78% |
| **UBI** *vs* UC | 0.980 | 0.869 | 0.203 | 0.203 | 0.00% | 2.90% |
| **UC** *vs* UBI | 0.974 | 0.869 | 0.252 | 0.083 | 0.36% | 3.99% |
| **UBI** *vs* UNIN | 0.973 | 0.838 | 0.234 | 0.125 | 0.72% | 2.17% |
| **UNIN** *vs* UBI | 0.955 | 0.838 | 0.308 | 0.208 | 1.09% | 4.71% |
| **UBI** *vs* UP | 0.984 | 0.904 | 0.183 | 0.042 | 11.96% | 0.72% |
| **UP** *vs* UBI | 0.972 | 0.904 | 0.254 | 0.333 | 0.72% | 18.84% |
| **UC** *vs* UNIN | 0.978 | 0.903 | 0.230 | 0.125 | 2.90% | 8.33% |
| **UNIN** *vs* UC | 0.973 | 0.903 | 0.238 | 0.167 | 1.81% | 5.07% |
| **UC** *vs* UP | 0.989 | 0.948 | 0.162 | 0.042 | 11.23% | 0.72% |
| **UP** *vs* UC | 0.986 | 0.948 | 0.184 | 0.208 | 0.36% | 14.13% |
| **UNIN** *vs* UP | 0.985 | 0.938 | 0.180 | 0.042 | 11.59% | 1.81% |
| **UP** *vs* UNIN | 0.984 | 0.938 | 0.195 | 0.125 | 0.00% | 16.67% |

## C. Comparison between Subjective Scores after PC and Mesh visualization

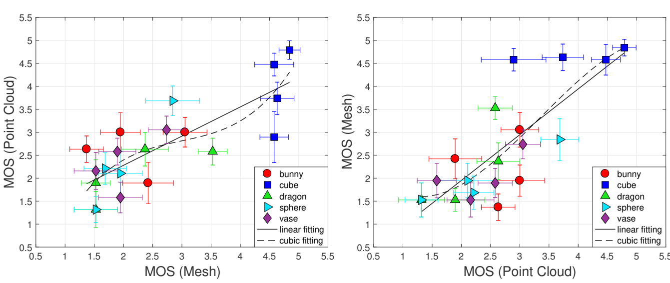
Finally, the subjective scores collected in this experiment were compared to ratings derived in a previous experiment, where the visual quality of the same degraded PCs was assessed without enabling any intervening reconstruction algorithm before rendering. The latter test was performed in the EPFL laboratory under identical conditions and using the same test equipment; a detailed description can be found in [7]. To avoid biases by the usage of different experimental

TABLE V
PERFORMANCE INDEXES TO COMPARE SUBJECTIVE SCORES AFTER PC AND MESH RENDERING (BOLD TEXT REPRESENTS THE GROUND TRUTH).

| | **PCC** | **SROCC** | **RMSE** | **OR** | **CE (%)** | **OE (%)** | **UE (%)** | **CD (%)** | **FR (%)** | **FD ( %)** | **FT (%)** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **PC** *vs* Mesh | 0.804 | 0.729 | 0.565 | 0.600 | 85% | 5% | 10% | 68.42% | 0% | 15.26% | 16.32% |
| **Mesh** *vs* PC | 0.808 | 0.729 | 0.702 | 0.500 | 80% | 10% | 10% | 68.95% | 0% | 5.79% | 25.26% |

settings (e.g., monitor), the statistical analysis is issued only on the EPFL scores. In particular, the performance indexes described in Section III-B are used to compare the two sets of scores. No, linear and cubic fitting functions were also tested, with the latter providing better fitting results that are used to report the performance indexes of Table V. In Figure 5, we provide scatter plots indicating the correlation between the two experiments, including every fitting function.

Based on our analysis, the correlation between these two tests is poor, indicating that the visual quality of identically distorted contents is affected by the use of an intervening surface reconstruction algorithm. To obtain a watertight object, commonly, the coordinates of the points are modified to best match the fitting surfaces. Hence, it is not straightforward whether the usage of a different 3D visual data representation, or the geometry errors introduced lead to this different rating trend; thus, further investigation is needed. The outcome of this analysis is that using a surface reconstruction technique as a pre-rendering step to consume 3D objects leads to differently rated visible distortions with respect to visualization of raw PC contents. To the best of our knowledge, this is the first study that confirms this statement through subjective testing.




(a) **PC** vs Mesh (b) **Mesh** vs PC

Fig. 5. No, Linear and Cubic fitting, to evaluate the correlation between PC and Mesh rendering (Bold text represents the ground truth).

## IV. CONCLUSIONS

In this work we conducted subjective evaluation of octree-based compression artifacts of PCs, rendered as mesh objects. The experiment was performed on five independent laboratories and our results reveal high correlation among test labs, although different displays were used. Comparison of the subjective scores of every lab with the state-of-the-art PC objective metrics shows that the visual quality cannot be sufficiently predicted for every type of content. Finally, a comparison between the ratings of subjects visualizing the same contents using two different visual data representations, namely sets of points and watertight surfaces, shows that the subjective results are affected by the usage of a surface reconstruction algorithm.

## ACKNOWLEDGMENTS

This work has been conducted in the framework of ImmersiaTV project under the European Union Horizon 2020 research and innovation program (grant agreement no. 688619) and funded by Swiss State Secretariat for Education, Research and Innovation SERI. The research was also supported by the FCT and FEDER PT2020 partnership agreement project PTDC/EEI-PRO/2849/ 2014 - POCI-01-0145-FEDER-016693, and FCT project UID/EEA/50008/2013. We would also like to thank George Paraskevopoulos and Christos Vassilopoulos for their efforts to conduct the experiment in UP.

## REFERENCES

[1] J. Zhang, W. Huang, X. Zhu, and J. N. Hwang, "A subjective quality evaluation for 3D point cloud models," in *2014 International Conference on Audio, Language and Image Processing*, July 2014, pp. 827–831.

[2] R. Mekuria, K. Blom, and P. Cesar, "Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, April 2017.

[3] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2017, pp. 1–6.

[4] M. Kazhdan and H. Hoppe, "Screened Poisson Surface Reconstruction," *ACM Trans. Graph.*, vol. 32, no. 3, pp. 29:1–29:13, July 2013.

[5] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2017, pp. 1–6.

[6] E. Alexiou and T. Ebrahimi, "On subjective and objective quality evaluation of point cloud geometry," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–3.

[7] E. Alexiou, E. Upenik, and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Proceedings of SPIE*, ser. Applications of Digital Image Processing XL, vol. 10396, Aug. 2017.

[8] ——, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2017, pp. 1–6.

[9] MPEG 3DG and Req., "Call for proposals for Point Cloud Compression V2," ISO/IEC MPEG2017/N16763, Hobart, AU, April 2017.

[10] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1–4.

[11] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Jan. 2012.

[12] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Evaluation Metrics for Point Cloud Compression," ISO/IEC MPEG2016/M39316, Chengdu, China, Oct. 2016.

[13] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul. 2012.

[14] ITU-T J.149, "Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)," International Telecommunication Union, Mar. 2004.